Appendix A

In this appendix, we prove Theorem 1 in the paper, which treats the case $e = c = 1$, with k and m arbitrary. For convenience, we reference equation numbers in the paper by preceding the equation number with a P. For example, to reference equation (16) in the paper, we write P(16) in this appendix.

We start by proving some preliminary lemmas.

Lemma 0 (Localization). *If A2 holds, then*

$$M(v) = (F_{V_1}(v_1), \ldots, F_{V_e}(v_e))$$

is a differentiable, invertible map from S_V onto S_U. Suppose, in addition, that A1 holds. Fix $z \in S_Z$. For each $x = (x_1, \ldots, x_e) \in S_{\mathcal{X}}$ there exists a unique $u \in S_U$ such that

$$u = M(v) = (I\!P\{\mathcal{X}_1 \le x_1 \mid Z = z\}, \ldots, I\!P\{\mathcal{X}_e \le x_e \mid Z = z\})\,.$$

Proof. The proof of the first part the lemma is trivial. We prove the second part for the case $e = 2$. Fix $v \in S_V$. By A1, for each $z \in S_Z$, there exists a unique $x = (x_1, x_2) \in S_{\mathcal{X}}$ such that $v = (v_1, v_2) = (\phi_1^{-1}(x_1, z), \phi_2^{-1}(x_2, z))$. By the first part of the lemma, there is a unique $u \in S_U$ for which $u = (u_1, u_2) = M(v) = (F_{V_1}(v_1), F_{V_2}(v_2))$. It follows that

$$\begin{aligned}
(u_1, u_2) &= (F_{V_1}(v_1), F_{V_2}(v_2)) \\
&= (F_{V_1}(\phi_1^{-1}(x_1, z)), F_{V_2}(\phi^{-1}(x_2, z))) \\
&= (I\!P\{V_1 \le \phi_1^{-1}(x_1, z)\}, I\!P\{V_2 \le \phi_2^{-1}(x_2, z)\}) \\
&= (I\!P\{V_1 \le \phi_1^{-1}(x_1, z) \mid Z = z\}, I\!P\{V_2 \le \phi_2^{-1}(x_2, z) \mid Z = z\}) \quad (V \text{ independent of } Z) \\
&= (I\!P\{\phi_1(z, V_1) \le x_1 \mid Z = z\}, I\!P\{\phi_2(z, V_2) \le x_2 \mid Z = z\}) \\
&= (I\!P\{\mathcal{X}_1 \le x_1 \mid Z = z\}, I\!P\{X_2 \le x_2 \mid Z = z\})\,.
\end{aligned}$$

This proves the second part of the Lemma. □

LEMMA 1 (Uniform Consistency). *If assumptions A1 through A19 hold, then*

$$\sup_{u \in U_\kappa} |\hat{\beta}(u) - \beta(u)| = o_p(1)\,.$$

as $n \to \infty$.

Proof. Fix $u \in U_\kappa$. Since $e = 1$, we estimate $\beta(u)$ with $\hat{\beta}(u) = \operatorname*{argmax}_{b \in B_u} \hat{S}_n(b \mid u)$ where

$$\hat{S}_n(b \mid u) = \frac{1}{n\tau_n} \sum_{j=1}^{n} (2Y_j - 1) K_n^*(X_j b) \mathcal{K}_n(\hat{U}_j - u) \tau_\kappa(\mathcal{Z}_j)\,.$$

Define

$$
\begin{aligned}
S_n(b \mid u) &= \frac{1}{n\tau_n}\sum_{j=1}^{n}(2Y_j-1)K_n^*(X_jb)\mathcal{K}_n(U_j-u)\tau_\kappa(\mathcal{Z}_j)\\
\bar{S}_n(b \mid u) &= \mathbb{E}S_n(b \mid u) = \frac{1}{\tau_n}\mathbb{E}(2Y-1)K_n^*(Xb)\mathcal{K}_n(U-u)\tau_\kappa(\mathcal{Z})\,.
\end{aligned}
$$

By iterated expectations, with the inner expectation over Y given U and Z, we get that

$$\bar{S}_n(b \mid u) = \frac{1}{\tau_n}\mathbb{E}g(U,Z)K_n^*(Xb)\mathcal{K}_n(U-u)\tau_\kappa(\mathcal{Z})$$

where the expectation $\mathbb{E}$ in the last expression is over U and Z and

$$g(U,Z) = 2\mathbb{P}(\epsilon(U) < X\beta(U) \mid U,Z) - 1 \tag{1}$$

where $\epsilon(U) = -X(B-\beta(U))$. By A1 and A2, $\mathcal{X} = \phi(Z, M^{-1}(U))$ and so given U and Z, $\mathcal{X}$ is determined. By A3, Z contains all the exogenous components of X. It follows that given U and Z, X is determined. Thus, conditioning on U and Z in (1) implies conditioning on U and X in (1). Deduce from P(16) that $g(u,Z) = 0$ whenever $X\beta(u) = 0$. This is a critical fact used in what follows.

Recall that $\mathcal{X} = \phi(Z, M^{-1}(U))$. We write $X \equiv X(U,Z)$ to acknowledge the functional dependence of X on U and Z. Apply a change of variable argument with $r = (U-u)/\tau_n$, to get

$$\bar{S}_n(b \mid u) = \mathbb{E}\left[\int g(u+r\tau_n, Z)K_n^*((X(u+r\tau_n,Z)b)f(u+r\tau_n)\mathcal{K}(r)dr\right]\tau_\kappa(\mathcal{Z})$$

where the expectation $\mathbb{E}$ is over Z and $f(\cdot)$ is the marginal density of U.

Next, define

$$
\begin{aligned}
\tilde{S}_n(b \mid u) &= \mathbb{E}\left[\int g(u+r\tau_n, Z)f(u+r\tau_n)\mathcal{K}(r)dr\right]\{X(u,Z)b>0\}\tau_\kappa(\mathcal{Z})\\
S(b \mid u) &= \mathbb{E}g(u,Z)f(u)\{X(u,Z)b>0\}\tau_\kappa(\mathcal{Z})\,.
\end{aligned}
$$

To show uniform consistency, we show that (i) a standard identification condition, called a strong maximization condition, holds uniformly over U_κ and (ii) as $n \to \infty$, $\hat{S}_n(b \mid u)$ converges in probability to $S(b \mid u)$ uniformly over the compact set $U_\kappa \otimes B_u$. (Pointwise strong maximization and a weak law of large numbers holding uniformly over B_u are sufficient to prove pointwise consistency while the stronger conditions (i) and (ii) are sufficient to prove uniform consistency.)

Start with the uniform strong maximization condition (i) above, which says that for any $\delta > 0$,

$$\inf_{u\in U_\kappa}\left[S(\beta(u) \mid u) - \sup_{|b-\beta(u)|\geq\delta} S(b \mid u)\right] > 0\,.$$

To establish this uniform condition, we first show the pointwise condition, namely, that the term in brackets above is positive for each $u \in U_\kappa$.[1]

Fix $u \in U_\kappa$. To prove the pointwise result, we show that $S(b \mid u)$ is continuous in b on the compact set B_u, and is uniquely maximized at $\beta(u)$. To see this, note that $S(b \mid u)$ is continuous in b by a dominated convergence argument using the almost sure continuity and uniform boundedness of the integrand $g(u, Z)f(u)\{X(u, Z)b > 0\}\tau_\kappa(\mathcal{Z})$. As before, A1, A2, and A3 imply that conditioning on Z and U in (1) implies conditioning on X and U in (1). Deduce from P(16) that for each $z \in S_Z$, the integrand attains its maximum value of $\max\{0, g(u, z)f(u)\tau_\kappa(z)\}$ at $b = \beta(u)$. It follows that $S(b \mid u)$ is maximized at $b = \beta(u)$. Unique maximization follows from P(16), A4, A5, and arguments in Horowitz (1998, pp.59-60). This establishes the pointwise result.

We now establish the uniform result. By A16, $g(u, z)f(u)$ is a continuous function of u. This and a dominated convergence argument similar to the one used to prove that $S(b \mid u)$ is continuous in b, imply that $S(b \mid u)$ is also continuous in u. Since $S(b \mid u)$ is continuous in both b and u, the maximum theorem implies that $\beta(u)$ is an UHC correspondence. This and the fact that $\beta(u)$ uniquely maximizes $S(b \mid u)$ imply that $\beta(u)$ is a continuous function. Since B_u is compact, the correspondence from u to B_u is compact-valued. These last two facts and A10 imply that the constraint set $\{b \in B_u : |b - \beta(u)| \geq \delta\}$ is a compact-valued UHC correspondence. This and continuity of $S(b \mid u)$ in both b and u imply that the constrained value function $\sup_{|b-\beta(u)|\geq\delta} S(b \mid u)$ is an USC function of u. Since the unconstrained value function $S(\beta(u) \mid u)$ is continuous in u, it follows that $S(\beta(u) \mid u) - \sup_{|b-\beta(u)|\geq\delta} S(b \mid u)$ is a LSC function of u. By the Weierstrass Theorem, this function must attain its minimum value on the compact set U_κ. By the pointwise result, this function is positive for each $u \in U_\kappa$. It follows that its minimized value must also be positive, which establishes condition (i). See Aliprantis and Border (1994) and Berge (1997) for references.

To show (ii), we note that

$$\begin{aligned} |\hat{S}_n(b \mid u) - S(b \mid u)| \;\; &\leq \;\; |\hat{S}_n(b \mid u) - S_n(b \mid u)| \\ &+ \;\; |S_n(b \mid u) - \bar{S}_n(b \mid u)| \\ &+ \;\; |\bar{S}_n(b \mid u) - \tilde{S}_n(b \mid u)| \\ &+ \;\; |\tilde{S}_n(b \mid u) - S(b \mid u)| \,. \end{aligned}$$

Consider the first term on the RHS of the last expression. An argument based on a Taylor expansion of each $\hat{U}_j$ about U_j (as in the proof of (14) in Lemma 2 below) shows that the first term on the RHS is $o_p(1)$ uniformly over $U_\kappa \otimes B_u$. Standard empirical process results (see, for example, Lemma 3A in Sherman (1994)) show that as $n \to \infty$, the second term on the RHS is $o_p(1)$ uniformly over $U_\kappa \otimes B_u$. A Taylor expansion of $g(u + r\tau_n, Z)f(u + r\tau_n)$ about u implies that the fourth term is also $o_p(1)$ uniformly over $U_\kappa \otimes B_u$.

We now turn to the third term, which requires a bit more work. Recall the definition of $\bar{S}_n(b \mid u)$

[1] It is possible to replace U_κ with S_U in the uniform strong maximization argument. But this would require that S_V be compact.

and that $\tau_n \ll \sigma_n$. Write b_0 for the component of b corresponding to $\mathcal{X}$. By a Taylor expansion about u, we get that

$$\begin{aligned} K^*(X(u + r\tau_n, Z)b) &= K^*(X(u, Z)b) + (r\tau_n/\sigma_n)b_0\mathcal{K}_n(X^*b) \\ &= K^*(X(u, Z)b) + o(1) \end{aligned}$$

as $n \to \infty$ uniformly over $r \in [-1, 1]$ and $(u, b) \in U_\kappa \otimes B_u$. Assumption A16 implies that $|g(u + r\tau_n, Z)f(u + r\tau_n)|$ is bounded. Deduce that for some $c > 0$,

$$|\bar{S}_n(b \mid u) - \tilde{S}_n(b \mid u)| \leq c\mathbb{E}\,|K^*(X(u, Z)b/\sigma_n) - \{(X(u, Z)b > 0\}| + o(1)$$

where the last expectation is over $X(u, Z)$.

Recall that $X = (X_1, \tilde{X})$. Also, recall that the coefficient of X_1 is unity. Let $W = Xb = X_1 + \tilde{X}\tilde{b}$ and consider the transformation $(X_1, \tilde{X}) \mapsto (W, \tilde{X})$. This transformation is 1-1 and onto and the Jacobian of the transformation is unity. Let $f(x_1 \mid \tilde{x}, u)$ denote the density of X_1 given $\tilde{X} = \tilde{x}$ and $U = u$. By A6, this density is continuous in x_1. It follows that the last expectation is equal to

$$\int_{\tilde{x}} \left[\int_w |K^*(w/\sigma_n) - \{w > 0\}| f(w - \tilde{x}\tilde{b} \mid \tilde{x}, u)\, dw \right] f(\tilde{x} \mid u)\, d\tilde{x}\,.$$

Note that this transformation shifts the dependence on u and b from the integrand, which depends on n, to the argument of the conditional density, which does not depend on n. Since B_u and U_κ are compact, A6 implies that there exist $b^* \in B_u$ and $u^* \in U_\kappa$ such that $f(w - \tilde{x}\tilde{b}^* \mid \tilde{x}, u^*) = \sup_{b \in B_u, u \in U_\kappa} f(w - \tilde{x}\tilde{b} \mid \tilde{x}, u)$. It follows that $f(w - \tilde{x}\tilde{b}^* \mid \tilde{x}, u^*)$ is a density with respect to lebesgue measure on $\mathbb{R}$, and the integral in brackets is bounded by the integral with $f(w - \tilde{x}\tilde{b} \mid \tilde{x}, u)$ replaced by $f(w - \tilde{x}\tilde{b}^* \mid \tilde{x}, u^*)$. For each fixed w, $|K^*(w/\sigma_n) - \{w > 0\}|$ is bounded and converges to zero as $n \to \infty$. By the DCT, the integral in brackets converges to zero as $n \to \infty$. Since U_κ is compact, A7 implies that there exists $u^* \in U_\kappa$ such that $f(\tilde{x} \mid u^*) = \sup_{u \in U_\kappa} f(\tilde{x} \mid u)$. A similar DCT argument shows that the outer integral also converges to zero as $n \to \infty$. Moreover, this convergence is uniform over $U_\kappa \otimes B_u$. This establishes (ii), proving Lemma 1. □

LEMMA 2 (Rates of Uniform Convergence). *If assumptions A1 through A19 hold, then*

$$\sup_{x \in \mathcal{X}_\kappa, z \in Z_\kappa} |\hat{U}(x, z) - U(x, z)| = O_p(1/\sqrt{n}\alpha_n)$$

as $n \to \infty$.

Proof. Recall that Z denotes the $1 \times m$ vector of instruments for the single endogenous regressor $\mathcal{X}$, and that $\mathcal{Z}$ denotes the single continuous instrument for $\mathcal{X}$. Also, recall that D denotes the $1 \times (m - 1)$ vector of discrete instruments for $\mathcal{X}$. We write $z = (z_0, d)$ for a typical point in the support of Z, where z_0 is a point in the support of $\mathcal{Z}$, and d is a point in the support of D. Fix $x \in \mathcal{X}_\kappa$

and $z = (z_0, d) \in Z_\kappa$. We have that

$$\hat{U}(x,z) = \frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}/\hat{f}(z_0,d) \tag{2}$$

where

$$\hat{f}(z_0,d) = \frac{1}{n\alpha_n}\sum_{k=1}^{n}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}\,.$$

Note that $\hat{f}(z_0,d)$ estimates $f(z_0,d) = f(z_0 \mid d)\mathbb{P}\{D = d\}$ where $f(z_0 \mid d)$ denotes the conditional density of $\mathcal{Z}$ given d evaluated at z_0. Abbreviate $\hat{f}(z_0,d)$ to $\hat{f}$ and $f(z_0,d)$ to f. Note that

$$\begin{aligned}
\hat{U}(x,z) &= \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}}{f(z_0,d)}\left[\frac{f}{\hat{f}}\right] \\
&= \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}}{f(z_0,d)}\left[1 - \left(1 - \frac{\hat{f}}{f}\right)\right]^{-1} \\
&= \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}}{f(z_0,d)}\left[1 + \left(1 - \frac{\hat{f}}{f}\right) + \left(1 - \frac{\hat{f}}{f}\right)^2 + \cdots\right].
\end{aligned}$$

We now analyze the leading term in this last expansion. Nonleading terms can be handled similarly and have smaller stochastic order. By a slight abuse of notation, we take

$$\hat{U}(x,z) = \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}}{f(z_0,d)}\,.$$

Write

$$\hat{U}(x,z) - U(x,z) = \hat{U}(x,z) - \mathbb{E}_z\hat{U}(x,z) + \mathbb{E}_z\hat{U}(x,z) - U(x,z)$$

where the expectation $\mathbb{E}_z$ is conditional on $Z = (z_0,d)$. By the first part of A15, $f(z_0 \mid d)$ is bounded above zero on Z_κ, precluding ratio bias. Since $x \in \mathcal{X}_\kappa$, $U(x,z)$ is eventually more than a bandwidth α_n from either boundary of S_U (0 or 1), precluding boundary bias. These facts, A14, the second part of A15, and a standard change of variable argument followed by a Taylor expansion to p_a terms implies that the bias term $\mathbb{E}_z\hat{U}(x,z) - U(x,z)$ has order $o_p(1/\sqrt{n})$ as $n \to \infty$, uniformly over $\mathcal{X}_\kappa \otimes Z_\kappa$. Note that

$$\hat{U}(x,z) - \mathbb{E}_z\hat{U}(x,z) = \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\} - \mathbb{E}_z\{\mathcal{X}_k \le x\}\mathcal{K}_n(\mathcal{Z}_k - z_0)\{D_k = d\}}{f(z_0,d)}\,. \tag{3}$$

This is a zero-mean empirical process. A1, A14, and standard empirical process results (see, for example, the proof of Lemma 3A in Sherman (1994)) imply that this last term has order $O_p(1/\sqrt{n}\alpha_n)$ as $n \to \infty$, uniformly over $\mathcal{X}_\kappa \otimes Z_\kappa$. This proves Lemma 2. $\square$

LEMMA 3 (Rates of Uniform Consistency). *If assumptions A1 through A19 hold, then*

$$\sup_{u\in U_\kappa} |\hat{\beta}(u) - \beta(u)| = O_p\left(1/\sqrt{n}\alpha_n\sigma_n\tau_n^2\right)$$

as $n \to \infty$.

Proof. Fix $u \in U_\kappa$. Recall $S_n(b \mid u) = \frac{1}{n\tau_n}\sum_{j=1}^n (2Y_j - 1)K_n^*(X_j b)\mathcal{K}_n(U_j - u)\tau_\kappa(\mathcal{Z}_j)$. Define $\bar{\beta}(u) = \underset{b\in B_u}{\operatorname{argmax}}\, S_n(b \mid u)$. Then

$$\hat{\beta}(u) - \beta(u) = [\hat{\beta}(u) - \bar{\beta}(u)] + [\bar{\beta}(u) - \beta(u)]\,. \tag{4}$$

Start with the second term on the RHS of (4). Define the gradient and hessian of $S_n(b \mid u)$:

$$\begin{aligned} G_n(b \mid u) &= \frac{1}{n\sigma_n\tau_n}\sum_{j=1}^n (2Y_j - 1)\mathcal{K}_n(X_j b)\tilde{X}_j'\mathcal{K}_n(U_j - u)\tau_\kappa(\mathcal{Z}_j) \\ H_n(b \mid u) &= \frac{1}{n\sigma_n^2\tau_n}\sum_{j=1}^n (2Y_j - 1)\mathcal{K}_n'(X_j b)\tilde{X}_j'\tilde{X}_j\mathcal{K}_n(U_j - u)\tau_\kappa(\mathcal{Z}_j)\,. \end{aligned}$$

The gradient and hessian of population criterion function $S(b \mid u)$ are denoted $G(b \mid u)$ and $H(b \mid u)$. By definition of $\bar{\beta}(u)$, $0 = G_n(\bar{\beta}(u) \mid u)$. A one term Taylor expansion of $G_n(\bar{\beta}(u) \mid u)$ about $\beta(u)$ implies that

$$\bar{\beta}(u) - \beta(u) = -[H_n(\bar{\beta}^*(u) \mid u)]^{-1} G_n(\beta(u) \mid u) \tag{5}$$

where $\bar{\beta}^*(u)$ is between $\bar{\beta}(u)$ and $\beta(u)$. Note that for each $u \in U_\kappa$,

$$\begin{aligned} H_n(\bar{\beta}^*(u) \mid u) &= H_n(\bar{\beta}^*(u) \mid u) - H(\bar{\beta}^*(u) \mid u) \\ &+ H(\bar{\beta}^*(u) \mid u) - H(\beta(u) \mid u)\,. \end{aligned}$$

The first term on the RHS of the last expression is bounded by

$$\sup_{(u,b)\in U_\kappa\otimes B_u} |H_n(b \mid u) - H(b \mid u)|\,.$$

The difference $H_n(b \mid u) - H(b \mid u)$ has mean zero for each $(u, b) \in U_\kappa \otimes B_u$. Standard empirical process arguments (once again, see Lemma 3A in Sherman (1994)) show that this last expression has order $O_p(1/\sqrt{n}\sigma_n^2\tau_n)$ as $n \to \infty$. Invoke A18. By a Taylor expansion of each of the k^2 components of $H(\bar{\beta}^*(u) \mid u)$ about $\beta(u)$, we get that the (i, j)th component of $H(\bar{\beta}^*(u) \mid u) - H(\beta(u) \mid u)$ equals $D_{ij}(\bar{\beta}^{**}(u) \mid u)(\bar{\beta}^*(u) - \beta(u))$, where $D_{ij}(b \mid u)$ is the partial derivative of the ijth component of $H(b \mid u)$ with respect to b, and $\bar{\beta}^{**}(u)$ is between $\bar{\beta}(u)$ and $\beta(u)$. By A18, $D_{ij}(b \mid u)$ is a continuous function on the compact set $B_u \otimes U_\kappa$. Thus, this term has order $o_p(1)$ uniformly over $u \in U_\kappa$ provided $\sup_{u\in U_\kappa} |\bar{\beta}(u) - \beta(u)| = o_p(1)$ as $n \to \infty$. But this uniformity result holds by arguments similar to (and simpler than) those used to prove Lemma 1. Provided $\sigma_n^2\tau_n \gg n^{-1/2}$, we get that uniformly over

$u \in U_\kappa$, as $n \to \infty$,

$$H_n(\bar{\beta}^*(u) \mid u) - H(\beta(u) \mid u) = O_p(1/\sqrt{n}\sigma_n^2\tau_n) + O_p(\sup_{u \in U_\kappa} |\bar{\beta}(u) - \beta(u)|) = o_p(1)$$

Now apply a Taylor expansion of $\left[H_n(\bar{\beta}^*(u) \mid u)\right]^{-1}$ about $H(\beta(u) \mid u)$. Provided $\sigma_n^2\tau_n \gg n^{-1/2}$, we get that uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\left[H_n(\bar{\beta}^*(u) \mid u)\right]^{-1} - \left[H(\beta(u) \mid u)\right]^{-1} = O_p(1/\sqrt{n}\sigma_n^2\tau_n) + O_p(\sup_{u \in U_\kappa} |\bar{\beta}(u) - \beta(u)|) = o_p(1)\,. \tag{6}$$

Further, note that A8, A18, A19, and continuity of the inverse function imply that uniformly over $u \in U_\kappa$,

$$[H(\beta(u) \mid u)]^{-1} = O(1)\,. \tag{7}$$

Deduce from (5), (6), and (7) that, uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\bar{\beta}(u) - \beta(u) = -\left[[H(\beta(u) \mid u)]^{-1} + o_p(1)\right] G_n(\beta(u) \mid u) = O_p(1) G_n(\beta(u) \mid u)\,. \tag{8}$$

We now turn to an analysis of $G_n(\beta(u) \mid u)$. We have that

$$G_n(\beta(u) \mid u) = [G_n(\beta(u) \mid u) - I\!\!E G_n(\beta(u) \mid u)] + I\!\!E G_n(\beta(u) \mid u)\,. \tag{9}$$

Note that the term in brackets is a zero-mean empirical process. Standard empirical process arguments show that, uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$G_n(\beta(u) \mid u) - I\!\!E G_n(\beta(u) \mid u) = O_p(1/\sqrt{n}\sigma_n\tau_n)\,. \tag{10}$$

We now show that the bias term $I\!\!E G_n(\beta(u) \mid u)$ can be neglected. That is, we show that, uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$I\!\!E G_n(\beta(u) \mid u) = o_p(1/\sqrt{n})\,. \tag{11}$$

Note that

$$I\!\!E G_n(\beta(u) \mid u) = \frac{1}{\sigma_n\tau_n} I\!\!E (2Y_j - 1)\mathcal{K}_n(X_j\beta(u))\tilde{X}_j' \mathcal{K}_n(U_j - u)\tau_\kappa(\mathcal{Z}_j)\,. \tag{12}$$

Holding u fixed, we will evaluate this expectation in four steps: (i) average over Y_j given U_j and Z_j (ii) average over U_j given Z_j and $X_j\beta(u)$ (iii) average over Z_j given $X_j\beta(u)$ and (iv) average over $X_j\beta(u)$.

Recall the definition of $g(U, Z)$ given in (1), as well as the key identification result in P(16) which follows from A11 and A12. After applying step (i), we get that the integrand in (12) equals

$$\frac{1}{\sigma_n\tau_n} g(U_j, Z_j)\mathcal{K}_n(X_j\beta(u))\tilde{X}_j' \mathcal{K}_n(U_j - u)\tau_\kappa(\mathcal{Z}_j)\,.$$

In applying step (ii), there are two cases to consider. The first is the case where $U_j = U_{j1}$. The second is the case where $U_j \neq U_{j1}$. We will analyze the former. The analysis of the latter is similar. Note that when $U_j = U_{j1}$, the random variable $\tilde{X}_j$ does not involve U_j. Apply step (ii), making the change of variable $r = (U_j - u)/\tau_n$. After step (ii), the integrand in (12) equals

$$\frac{1}{\sigma_n}\left[\int g(u+\tau_n r, Z_j) f(u+\tau_n r \mid Z_j, X_j\beta(u))\mathcal{K}(r)dr\right]\mathcal{K}_n(X_j\beta(u))\tilde{X}_j'\tau_\kappa(\mathcal{Z}_j)$$

where $f(\cdot \mid Z, X\beta(u))$ denotes the density of U given Z and $X\beta(u)$.

In applying step (iii), write $\Gamma_n(X_j\beta(u))$ for the expectation over Z_j given $X_j\beta(u)$ of

$$\left[\int g(u+\tau_n r, Z_j) f(u+\tau_n r \mid Z_j, X_j\beta(u))\mathcal{K}(r)dr\right]\tilde{X}_j'\tau_\kappa(\mathcal{Z}_j)\,.$$

After applying step (iii), the integrand in (12) equals

$$\frac{1}{\sigma_n}\mathcal{K}_n(X_j\beta(u))\Gamma_n(X_j\beta(u))\,.$$

Finally, apply step (iv), making the change of variable $s = X_j\beta(u)/\sigma_n$ to get that the bias term in (12) equals

$$\int \Gamma_n(\sigma_n s) f(\sigma_n s)\mathcal{K}(s)ds$$

where $f(\cdot)$ denotes the density of $X_j\beta(u)$. Apply assumptions A14 and A16, and expand the product $g(u+\tau_n r, Z_j) f(u+\tau_n r \mid Z_j, X_j\beta(u))$ about $U_j = u$ to p_τ terms to replace it with $g(u, Z_j) f(u \mid Z_j, X_j\beta(u))$ plus a term that is $o_p(1/\sqrt{n})$ as $n \to \infty$. Then, apply A14 and A17 and expand $\Gamma_n(\sigma_n s) f(\sigma_n s)$ about $X_j\beta(u) = 0$ to p_σ terms to replace $\Gamma_n(\sigma_n s) f(\sigma_n s)$ with zero plus a term that is $o_p(1/\sqrt{n})$ as $n \to \infty$. The leading term in this expansion is zero because $g(u, Z_j) = 0$ when $X_j\beta(u) = 0$. The latter follows from P(16). This proves (11).

It follows from (8), (9), (10), and (11) that, as $n \to \infty$,

$$\sup_{u\in U_\kappa} |\bar{\beta}(u) - \beta(u)| = O_p(1/\sqrt{n}\sigma_n\tau_n)\,. \tag{13}$$

Next we show that, as $n \to \infty$,

$$\sup_{u\in U_\kappa} |\hat{\beta}(u) - \bar{\beta}(u)| = O_p(1/\sqrt{n}\alpha_n\sigma_n\tau_n^2)\,. \tag{14}$$

Define the gradient and hessian of $\hat{S}_n(b \mid u)$:

$$\begin{aligned}
\hat{G}_n(b \mid u) &= \frac{1}{n\sigma_n\tau_n}\sum_{j=1}^n (2Y_j - 1)\mathcal{K}_n(X_j b)\tilde{X}_j'\mathcal{K}_n(\hat{U}_j - u)\tau_\kappa(\mathcal{Z}_j) \\
\hat{H}_n(b \mid u) &= \frac{1}{n\sigma_n^2\tau_n}\sum_{j=1}^n (2Y_j - 1)\mathcal{K}_n'(X_j b)\tilde{X}_j'\tilde{X}_j\mathcal{K}_n(\hat{U}_j - u)\tau_\kappa(\mathcal{Z}_j)\,.
\end{aligned}$$

By definition of $\hat{\beta}(u)$, $0 = \hat{G}_n(\hat{\beta}(u) \mid u)$. A one term Taylor expansion of $\hat{G}_n(\hat{\beta}(u) \mid u)$ about $\beta(u)$

implies that

$$\hat{\beta}(u) - \beta(u) = -[\hat{H}_n(\hat{\beta}^*(u) \mid u)]^{-1}\hat{G}_n(\beta(u) \mid u) \tag{15}$$

where $\hat{\beta}^*(u)$ is between $\hat{\beta}(u)$ and $\beta(u)$. Deduce from (5) and (15) that

$$\hat{\beta}(u) - \bar{\beta}(u) = -\left[\hat{H}_n(\hat{\beta}^*(u) \mid u)\right]^{-1}\hat{G}_n(\beta(u) \mid u) + \left[H_n(\bar{\beta}^*(u) \mid u)\right]^{-1}G_n(\beta(u) \mid u)\,. \tag{16}$$

Note that

$$\hat{G}_n(\beta(u) \mid u) = \frac{1}{n\sigma_n\tau_n}\sum_{j=1}^{n}(2Y_j - 1)\mathcal{K}_n(X_j\beta(u))\tilde{X}_j'\mathcal{K}_n(\hat{U}_j - u)\tau_\kappa(\mathcal{Z}_j)\,.$$

If we Taylor expand each summand about U_j, then the sum of the first terms in these expansions equals $G_n(\beta(u) \mid u)$, a useful quantity to isolate in the subsequent analysis. By applying these expansions we get

$$\hat{G}_n(\beta(u) \mid u) = G_n(\beta(u) \mid u) + \frac{1}{n\sigma_n\tau_n}\sum_{j=1}^{n}\Lambda_n(\hat{U}_j^*, u)(\hat{U}_j - U_j)\tau_\kappa(\mathcal{Z}_j) \tag{17}$$

where $\hat{U}_j^*$ is between $\hat{U}_j$ and U_j, and

$$\begin{aligned}\Lambda_n(U,u) &= \frac{\partial}{\partial U}\left[(2Y-1)\mathcal{K}_n(X\beta(u))\tilde{X}'\mathcal{K}_n((U-u)\right] \\ &= (2Y-1)\mathcal{K}_n(X\beta(u))\tilde{X}'\mathcal{K}_n'((U-u)/\tau_n\,.\end{aligned}$$

Deduce from A14 that $\Lambda_n(U,u)\tau_\kappa(\mathcal{Z}) = O(1/\tau_n)$ as $n \to \infty$. Then apply Lemma 2 and (17) to get that, uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\hat{G}_n(\beta(u) \mid u) = G_n(\beta(u) \mid u) + O_p(1/\sqrt{n}\alpha_n\sigma_n\tau_n^2)\,. \tag{18}$$

Note that (10) and (11) imply that, uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$G_n(\beta(u) \mid u) = O_p(1/\sqrt{n}\sigma_n\tau_n)\,. \tag{19}$$

Now, consider the term $\hat{H}_n(\hat{\beta}^*(u) \mid u)$ in (15). We have that

$$\begin{aligned}\hat{H}_n(\hat{\beta}^*(u) \mid u) - H(\beta(u) \mid u) &= \hat{H}_n(\hat{\beta}^*(u) \mid u) - H_n(\hat{\beta}^*(u) \mid u) \\ &+ H_n(\hat{\beta}^*(u) \mid u) - H(\hat{\beta}^*(u) \mid u) \\ &+ H(\hat{\beta}^*(u) \mid u) - H(\beta(u) \mid u)\,.\end{aligned}$$

By arguments very similar to those used to establish (18), we get that the first term in the decomposition, uniformly over $u \in U_\kappa$, has order $O_p(1/\sqrt{n}\alpha_n\sigma_n^2\tau_n^2)$ as $n \to \infty$. Arguments made previously show that the second term in the decomposition, uniformly over $u \in U_\kappa$, has order $O_p(1/\sqrt{n}\sigma_n^2\tau_n)$ as $n \to \infty$, while the third term, uniformly over $u \in U_\kappa$, has order $O_p(\sup_{u\in U_\kappa}|\hat{\beta}(u) - \beta(u)|) = o_p(1)$ as $n \to \infty$. Then the Taylor expansion arguments used to establish (6) can be used to show that uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\begin{aligned}\left[\hat{H}_n(\hat{\beta}^*(u) \mid u)\right]^{-1} - [H(\beta(u) \mid u)]^{-1} &= O_p(1/\sqrt{n}\alpha_n\sigma_n^2\tau_n^2) + O_p(1/\sqrt{n}\sigma_n^2\tau_n) + O_p(\sup_{u \in U_\kappa} |\hat{\beta}(u) - \beta(u)|) \\ &= o_p(1)\,. \end{aligned} \tag{20}$$

Recall (16). Deduce from (7), (20), and (18), and then (7), (6), and (19), that (14) holds. Lemma 3 now follows from (4), (14), and (13). □

We are now in a position to prove that $\hat{\beta}_\kappa$ is a $\sqrt{n}$-consistent and asymptotically normally distributed estimator of β_κ.

LEMMA 4 (THE SECOND TERM IN P(10)). *If A1 through A19 hold, then*

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\beta}(U_i) - \beta(U_i)\right]\tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i) = \frac{1}{n}\sum_{i=1}^{n} f_n^{(2)}(W_i) + \frac{1}{n}\sum_{i=1}^{n} f_n^{(3)}(W_i) + o_p(1/\sqrt{n})$$

as $n \to \infty$, *where* $f_n^{(2)}(W_i) = f_n(P,P,W_i) + f_n(P,P,P,W_i)$ *and* $f_n^{(3)}(W_i) = f_n(P,W_i)$, *with* $f_n(W_i,W_j,W_k)$, $f_n(W_i,W_j,W_k,W_l)$, *and* $f_n(W_i,W_j)$ *defined in P(17), P(18), and P(19), respectively.*

PROOF. Consider the second term in P(10). This term equals

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\beta}(U_i) - \beta(U_i)\right]\tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i)\,. \tag{21}$$

For ease of notation, we suppress the trimming function $\tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i)$. We get

$$\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\beta}(U_i) - \bar{\beta}(U_i)\right] + \frac{1}{n}\sum_{i=1}^{n}\left[\bar{\beta}(U_i) - \beta(U_i)\right]\,. \tag{22}$$

Start with the first term in (22). By (16), this term equals

$$-\frac{1}{n}\sum_{i=1}^{n}\left[\left[\hat{H}_n(\hat{\beta}^*(U_i) \mid U_i)\right]^{-1}\hat{G}_n(\beta(U_i) \mid U_i) - \left[H_n(\bar{\beta}^*(U_i) \mid U_i)\right]^{-1} G_n(\beta(U_i) \mid U_i)\right]\,. \tag{23}$$

By (20) and Lemma 3 we get that uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\left[\hat{H}_n(\hat{\beta}^*(u) \mid u)\right]^{-1} = [H(\beta(u) \mid u)]^{-1} + O_p(1/\sqrt{n}\alpha_n\sigma_n^2\tau_n^2)\,. \tag{24}$$

By (6) and (13) we get that uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\left[H_n(\bar{\beta}^*(u) \mid u)\right]^{-1} = [H(\beta(u) \mid u)]^{-1} + O_p(1/\sqrt{n}\sigma_n^2\tau_n)\,. \tag{25}$$

By (18) and (19), we get that uniformly over $u \in U_\kappa$, as $n \to \infty$,

$$\hat{G}_n(\beta(u) \mid u) = O_p(1/\sqrt{n}\alpha_n\sigma_n\tau_n^2)\,. \tag{26}$$

Equations (24) and (26), together with (25) and (19), imply that the expression in (23) equals

$$\frac{1}{n}\sum_{i=1}^{n}\Big[[H(\beta(U_i)\mid U_i)]^{-1}\Big[\hat{G}_n(\beta(U_i)\mid U_i)-G_n(\beta(U_i)\mid U_i)\Big]\Big]+O_p(1/n\alpha_n^2\sigma_n^3\tau_n^4)\,. \tag{27}$$

Note that the $O_p(1/n\alpha_n^2\sigma_n^3\tau_n^4)$ term has order $o_p(1/\sqrt{n})$ provided $\alpha_n^2\sigma_n^3\tau_n^4 \gg n^{-1/2}$. By (17),

$$\hat{G}_n(\beta(U_i)\mid U_i)-G_n(\beta(U_i)\mid U_i)=\frac{1}{n\sigma_n\tau_n}\sum_{j=1}^{n}\Lambda_n(\hat{U}_j^*,U_i)(\hat{U}_j-U_j)\tau_\kappa(\mathcal{Z}_j)\,.$$

Lemma 2 and a Taylor expansion of $\Lambda_n(\hat{U}_j^*,U_i)$ about U_j (see the expression following (17)) imply that, uniformly over i and j, as $n\to\infty$,

$$\hat{G}_n(\beta(U_i)\mid U_i)-G_n(\beta(U_i)\mid U_i)=\frac{1}{n\sigma_n\tau_n}\sum_{j=1}^{n}\Lambda_n(U_j,U_i)(\hat{U}_j-U_j)\tau_\kappa(\mathcal{Z}_j)+O_p(1/n\alpha_n^2\sigma_n\tau_n^3)\,. \tag{28}$$

Note that the $O_p(1/n\alpha_n^2\sigma_n\tau_n^3)$ term has order $o_p(1/\sqrt{n})$ provided $\alpha_n^2\sigma_n\tau_n^3 \gg n^{-1/2}$.

As in the proof of Lemma 2, we have that

$$\hat{U}_j=\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)\{D_k=D_j\}/\hat{f}(\mathcal{Z}_j,D_j) \tag{29}$$

where

$$\hat{f}(\mathcal{Z}_j,D_j)=\frac{1}{n\alpha_n}\sum_{k=1}^{n}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)\{D_k=D_j\}\,.$$

Note that $\hat{f}(\mathcal{Z}_j,D_j)$ estimates $f(\mathcal{Z}_j,D_j)=f(\mathcal{Z}_j\mid D_j)I\!P\{D=D_j\}$ where $f(\mathcal{Z}_j\mid D_j)$ denotes the conditional density of $\mathcal{Z}_j$ given D_j. Abbreviate $\hat{f}(\mathcal{Z}_j,D_j)$ to $\hat{f}$ and $f(\mathcal{Z}_j,D_j)$ to f. Then

$$\begin{aligned}\hat{U}_j &= \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)\{D_k=D_j\}}{f(\mathcal{Z}_j,D_j)}\left[\frac{f}{\hat{f}}\right]\\ &= \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)\{D_k=D_j\}}{f(\mathcal{Z}_j,D_j)}\left[1-\left(1-\frac{\hat{f}}{f}\right)\right]^{-1}\\ &= \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)\{D_k=D_j\}}{f(\mathcal{Z}_j,D_j)}\left[1+\left(1-\frac{\hat{f}}{f}\right)+\left(1-\frac{\hat{f}}{f}\right)^2+\cdots\right]\,.\end{aligned} \tag{30}$$

The first two terms in the last expansion, when combined with (28) and (27), make first order asymptotic contributions. The remaining terms lead to contributions of order $o_p(1/\sqrt{n})$ and so can be neglected.[2]

[2]The first two terms in the last expansion, apart from a $o_p(1/\sqrt{n})$ bias term, are zero-mean U-statistics of orders three and four, respectively. Each of these U-statistics has a nondegenerate projection (the first term in the Hoeffding decomposition), resulting in a first order asymptotic contribution. We demonstrate this fact with the first term in the expansion. However, this does not happen with the higher order terms in the expansion. Take the third term, for example. Apart from a $o_p(1/\sqrt{n})$ bias term, this term is a zero-mean U-statistic of order five. It is straightforward to show that the average of its kernel function over either of two arguments, conditional on the remaining four arguments,

We now analyze the leading term in this last expansion. Analysis of the second term is very similar and so is omitted. By a slight abuse of notation, take

$$\hat{U}_j = \frac{\frac{1}{n\alpha_n}\sum_{k=1}^{n}\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\}}{f(\mathcal{Z}_j, D_j)} .$$

Write

$$\hat{U}_j - U_j = \hat{U}_j - \mathbb{E}_j\hat{U}_j + \mathbb{E}_j\hat{U}_j - U_j$$

where the expectation $\mathbb{E}_j$ is conditional on $(\mathcal{Z}_j, D_j)$. Invoke A13, A14, and A15 and apply a change of variable followed by a Taylor expansion to p_a terms to show that the bias term $\mathbb{E}_j\hat{U}_j - U_j$ has order $o_p(1/\sqrt{n})$. Therefore, it is enough to analyze

$$\hat{U}_j - \mathbb{E}_j\hat{U}_j = \frac{1}{n\alpha_n}\sum_{k=1}^{n}\frac{\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\} - \mathbb{E}_j\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\}}{f(\mathcal{Z}_j, D_j)} . \tag{31}$$

Substitute (31) for $\hat{U}_j - U_j$ in (28), then combine with (27) and expand sums to get

$$\frac{1}{n^3}\sum_{i,j,k}[H(\beta(U_i) \mid U_i)]^{-1}\Lambda_n(U_j, U_i)\frac{\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\} - \mathbb{E}_j\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\}}{\alpha_n\sigma_n\tau_n f(\mathcal{Z}_j, D_j)} \tag{32}$$

where, to save space we suppress the trimming function $\tau_\kappa(\mathcal{Z}_j)$. Note that there are n^3 summands in (32). Define $n_{(3)} = n(n-1)(n-2)$ and $\boldsymbol{i_3} = (i, j, k)$ where $i \ne j \ne k \ne i$. Then the term in (32) equals

$$\frac{1}{n_{(3)}}\sum_{\boldsymbol{i_3}}[H(\beta(U_i) \mid U_i)]^{-1}\Lambda_n(U_j, U_i)\frac{\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\} - \mathbb{E}_j\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\}}{\alpha_n\sigma_n\tau_n f(\mathcal{Z}_j, D_j)} \tag{33}$$

plus a term that can be neglected asymptotically. The reason is that there are only $O(n^2)$ terms in the difference between the triple sums in (32) and (33). If $\alpha_n\sigma_n\tau_n \gg n^{-1/2}$, then the difference between (32) and (33) has order $o_p(1/\sqrt{n})$ as $n \to \infty$.

The term in (33) is a zero-mean U-statistic of order three. Define $W_i = (Y_i, X_i, Z_i, U_i)$ and $f_n(W_i, W_j, W_k)$ to be the (i, j, k)th summand in expression (33). Define $n_{(2)} = n(n-1)$ and $\boldsymbol{i_2} = (i, j)$ where $i \ne j$. Apply the Hoeffding decomposition (see Serfling, 1980,Chapter 5) to get that

$$\begin{aligned}\frac{1}{n_{(3)}}\sum_{\boldsymbol{i_3}} f_n(W_i, W_j, W_k) &= \frac{1}{n}\sum_{i=1}^{n}[f_n(W_i, P, P) + f_n(P, W_i, P) + f_n(P, P, W_i)] \\ &+ \frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}} g_n(W_i, W_j) + \frac{1}{n_{(3)}}\sum_{\boldsymbol{i_3}} h_n(W_i, W_j, W_k)\end{aligned} \tag{34}$$

where the second average in the decomposition is a degenerate U-statistic of order two, and the third

is zero. This implies a zero projection, resulting in no first order asymptotic contribution. Moreover, the tail process is easily shown to be $o_p(1/\sqrt{n})$.

average is a degenerate U-statistic of order three. It follows that as $n \to \infty$, the second and third averages have order $O_p(1/n\alpha_n\sigma_n\tau_n)$ and $O_p(1/n^{3/2}\alpha_n\sigma_n\tau_n)$, respectively. Thus, if $\alpha_n\sigma_n\tau_n \gg n^{-1/2}$, then both of these terms have order $o_p(1/\sqrt{n})$ and so can be ignored.

We now show that the first term in (34) is $\sqrt{n}$-consistent and asymptotically normally distributed. First note that $f_n(W_i, P, P) = f_n(P, W_i, P) = 0$. To see this, fix W_i and W_j and note that $f_n(W_i, W_j, P) = 0$. So, it suffices to analyze the average of the $f_n(P, P, W_i)$'s in (34). For convenience, we will write this term as

$$\frac{1}{n}\sum_{k=1}^{n} f_n(P, P, W_k)\,. \tag{35}$$

The claim that this term is $\sqrt{n}$-consistent might initially be viewed with some suspicion. To see why, note that

$$\begin{aligned} f_n(W_i, W_j, W_k) \;&=\; \frac{1}{\alpha_n\sigma_n\tau_n^2} H(\beta(U_i) \mid U_i)]^{-1} \\ &\times\; (2Y_j - 1)\mathcal{K}_n(X_j\beta(U_i))\tilde{X}_j'\mathcal{K}_n'(U_j - U_i) \\ &\times\; \frac{\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\} - \mathbb{E}_j\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)\{D_k = D_j\}}{f(\mathcal{Z}_j, D_j)}\,. \end{aligned}$$

We see that $f_n(W_i, W_j, W_k)$ is a product of terms divided by $\alpha_n\sigma_n\tau_n^2$. However, this product involves only three kernel function factors: the kernel function used to estimate U corresponding to α_n, the kernel function used to smooth the indicator $\{t > 0\}$ corresponding to σ_n, and the derivative of the kernel function used to localize on U corresponding to τ_n. Integrating a kernel function or its derivative involves a change of variable, resulting in a rescaling by the corresponding bandwidth factor. Thus, one might expect that the one α_n factor, the one σ_n factor, and one of the τ_n factors can be accounted for, but not the remaining τ_n factor. If true, this would imply that the expression in (35) is at best $\sqrt{n\tau_n}$-consistent, but not $\sqrt{n}$-consistent. But, in fact, the expression in (35) is $\sqrt{n}$-consistent. The reason is that the derivative of the bias reducing kernel we use is an odd function, which, when integrated, annihilates a leading constant term, thus accounting for the fourth bandwidth factor. We now show this.

To save space, we will consider the case $m = 1$ so that $Z_j = \mathcal{Z}_j$. The case of general m adds nothing to understanding and follows immediately from the argument given below by replacing marginal densities with joint densities (products of conditional and marginal densities) and adding summations over the discrete conditioning variables. From (33) and the expression following (17), we get that the term in question equals

$$\frac{1}{n}\sum_{k=1}^{n} f_n^{(1)}(P, P, W_k) \tag{36}$$

where $f_n^{(1)}(P,P,W_k)$ equals

$$\frac{1}{\alpha_n\sigma_n\tau_n^2}\mathbb{E}_k h(U_i)(2Y_j-1)\mathcal{K}_n(X_j\beta(U_i))\tilde{X}_j'\mathcal{K}_n'(U_j-U_i)\frac{\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)-\mathbb{E}_j\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)}{f(\mathcal{Z}_j)} \tag{37}$$

where the expectation $\mathbb{E}_k$ is over W_i and W_j given W_k, $h(U_i)=[H(\beta(U_i)\mid U_i)]^{-1}$, and the expectation $\mathbb{E}_j$ is the expectation over W_k given W_j.

In evaluating the expectation $\mathbb{E}_k$ in (37), we first fix W_i and average over W_j. Note that the integrand depends on W_i only through U_i. For ease of notation, when averaging out over W_j, we will replace each U_i with u. The averaging over W_j will be done in four steps: (i) average over Y_j given U_j and $\mathcal{Z}_j$ (ii) average over U_j given $\mathcal{Z}_j$ and $X_j\beta(u)$ (iii) average over $\mathcal{Z}_j$ given $X_j\beta(u)$ and (iv) average over $X_j\beta(u)$. After step (iv), we will average over u to get $f_n^{(1)}(P,P,W_k)$.

Recall the definition of $g(\cdot,\cdot)$ in (1). After applying step (i) the integrand in (37) equals

$$\frac{1}{\alpha_n\sigma_n\tau_n^2}h(u)g(U_j,Z_j)\mathcal{K}_n(X_j\beta(u))\tilde{X}_j'\mathcal{K}_n'(U_j-u)\frac{\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)-\mathbb{E}_j\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)}{f(\mathcal{Z}_j)}\,. \tag{38}$$

In applying step (ii), there are two cases to consider, namely, the case $U_j=U_{1j}$ and the case $U_j\neq U_{1j}$. We analyze the former case. Analysis of the latter case is similar. Note that when $U_j=U_{1j}$, then $\tilde{X}_j$ does not depend on U_j. Make the change of variable $r=(U_j-u)/\tau_n$. After applying step (ii), the integrand in (37) equals

$$\begin{aligned}\frac{1}{\alpha_n\sigma_n\tau_n}\mathcal{K}_n(X_j\beta(u)) \quad&\times\quad h(u)\tilde{X}_j'\frac{\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)-\mathbb{E}_j\{\mathcal{X}_k\le\mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k-\mathcal{Z}_j)}{f(\mathcal{Z}_j)}\\ &\times\quad\int g(u+\tau_n r,Z_j)f(u+\tau_n r\mid\mathcal{Z}_j,X_j\beta(u))\mathcal{K}'(r)dr\,. \end{aligned}\tag{39}$$

where $f(\cdot\mid\mathcal{Z},X\beta(u))$ is the conditional density of U given $\mathcal{Z}$ and $X\beta(u)$. We now closely examine the integral in (39). By Taylor expansions about u we get that

$$\begin{aligned} g(u+\tau_n r,Z_j) &= g(u,Z_j)+\tau_n r g_1(u^*,Z_j)\\ f(u+\tau_n r\mid\mathcal{Z}_j,X_j\beta(u)) &= f(u\mid\mathcal{Z}_j,X_j\beta(u))+\tau_n r f_1(\bar{u}\mid\mathcal{Z}_j,X_j\beta(u)) \end{aligned}$$

where g_1 is the partial derivative of g with respect to its first argument, and u^* is between u and $u+\tau_n r$, while $f_1(\cdot\mid\mathcal{Z}_j)$ is the partial derivative of $f(\cdot\mid\mathcal{Z}_j,X_j\beta(u))$ with respect to its first argument, and $\bar{u}$ is between u and $u+\tau_n r$. Since $\mathcal{K}'(\cdot)$ is an odd function integrated over a symmetric interval, the integral of the leading constant term is annihilated:

$$\int g(u,Z_j)f(u\mid\mathcal{Z}_j,X_j\beta(u))\mathcal{K}'(r)dr=g(u,Z_j)f(u\mid\mathcal{Z}_j,X_j\beta(u))\int\mathcal{K}'(r)dr=0\,. \tag{40}$$

Deduce that the integral in (39) equals

$$\tau_n \int r\left[g(u, Z_j) f_1(\bar{u} \mid \mathcal{Z}_j, X_j\beta(u)) + f(u \mid \mathcal{Z}_j, X_j\beta(u)) g_1(u^*, Z_j) + r\tau_n f_1(\bar{u} \mid \mathcal{Z}_j, X_j\beta(u)) g_1(u^*, Z_j)\right] \mathcal{K}'(r) dr\,. \tag{41}$$

Let $I_n(\mathcal{Z}_j, X_j\beta(u), u)$ denote the integral in (41). Thus, after applying step (ii), the integrand in (37) equals

$$\frac{1}{\alpha_n\sigma_n}\mathcal{K}_n(X_j\beta(u))h(u)\tilde{X}_j' \frac{[\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j) - \mathbb{E}_j\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)]}{f(\mathcal{Z}_j)} I_n(\mathcal{Z}_j, X_j\beta(u), u)\,. \tag{42}$$

We see that in applying step (ii), the two τ_n factors have been accounted for. Define $I_n(\mathcal{Z}_j) = \mathbb{E}_j\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}_n(\mathcal{Z}_k - \mathcal{Z}_j)$. Apply step (iii). Make the change of variable $s = (\mathcal{Z}_j - \mathcal{Z}_k)/\alpha_n$. After applying step (iii) the integrand in (37) equals

$$\frac{1}{\sigma_n}\mathcal{K}_n(X_j\beta(u)) \int h(u)\tilde{X}_j' I_n(\mathcal{Z}_k + \tau_n s, X_j\beta(u), u) \frac{\{\mathcal{X}_k \le \mathcal{X}_j\}\mathcal{K}(s) - I_n(\mathcal{Z}_k + \tau_n s)}{f(\mathcal{Z}_k + \tau_n s)} f(\mathcal{Z}_k + \tau_n s \mid X_j\beta(u)) ds\,. \tag{43}$$

where $f(\cdot \mid X_j\beta(u))$ is the conditional density of $\mathcal{Z}$ given $X\beta(u) = X_j\beta(u)$. Let $I_n(W_k, X_j\beta(u), u)$ denote the integral in (43). Then the integrand in (37) equals

$$\frac{1}{\sigma_n}\mathcal{K}_n(X_j\beta(u)) I_n(W_k, X_j\beta(u), u)\,. \tag{44}$$

We now apply step (iv). Make the change of variable $t = X_j\beta(u)/\sigma_n$. After applying step (iv) the integrand in (37) equals

$$\int I_n(W_k, \sigma_n t, u) f(\sigma_n t \mid u)\mathcal{K}(t) dt \tag{45}$$

where $f(\cdot \mid u)$ is the density of $X_j\beta(u)$. Finally, we average out over u to get that the expression in (37) equals

$$\int \left[\int I_n(W_k, \sigma_n t, u) f(\sigma_n t \mid u)\mathcal{K}(t) dt\right] f(u) du \tag{46}$$

where $f(\cdot)$ denotes the marginal density of U. That is, $f_n^{(1)}(P, P, W_k)$ in (36) is equal to this last expression. We see that the expression in (36) is an average of zero mean iid random vectors. Moreover, because all components of $f_n^{(1)}$ are bounded and the density of $\mathcal{Z}$ is bounded away from zero on $\{|\mathcal{Z}| \le \kappa\}$, these variables have finite variance as well. Deduce from a standard CLT that the expression in (36) is $\sqrt{n}$-consistent and asymptotically normally distributed. This takes care of the first term in (22).

Now we analyze the second term in (22). This term is much easier to analyze than the first term in (22) because it does not involve estimated U_i's. By (5), we get that

$$\frac{1}{n}\sum_{i=1}^{n}\left[\bar{\beta}(U_i) - \beta(U_i)\right] = \frac{1}{n}\sum_{i=1}^{n}\left[H_n(\bar{\beta}^*(U_i) \mid U_i)\right]^{-1} G_n(\beta(U_i) \mid U_i)\,. \tag{47}$$

By (6), (13), and (19), we get that as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[H_n(\bar{\beta}^*(U_i) \mid U_i)\right]^{-1} G_n(\beta(U_i) \mid U_i) = \frac{1}{n}\sum_{i=1}^{n}\left[H(\beta(U_i) \mid U_i)\right]^{-1} G_n(\beta(U_i) \mid U_i) + O_p(1/n\sigma_n^3\tau_n^2)\,. \tag{48}$$

Provided $\sigma_n^3\tau_n^2 \gg n^{-1/2}$, the $O_p(1/n\sigma_n^3\tau_n^2)$ term has order $o_p(1/\sqrt{n})$. Consider the main term on the RHS of equation (48). Recall that $W_i = (Y_i, X_i, Z_i, U_i)$ and $h(u) = [H(\beta(u) \mid u)]^{-1}$. Define

$$f_n(W_i, W_j) = \frac{1}{\sigma_n\tau_n} h(U_i)(2Y_j - 1)\mathcal{K}_n(X_j\beta(U_i))\tilde{X}_j'\mathcal{K}_n(U_j - U_i)\,. \tag{49}$$

where, as before, to save space we have suppressed the trimming function $\tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i)\tau_\kappa(\mathcal{Z}_j)$. We get that the main term on the RHS of (48) is equal to

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} f_n(W_i, W_j)\,. \tag{50}$$

There are only n terms in the double sum for which $i = j$. Provided $\sigma_n\tau_n \gg n^{-1/2}$, as $n \to \infty$,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} f_n(W_i, W_j) = \frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}} f_n(W_i, W_j) + o_p(1/\sqrt{n})\,. \tag{51}$$

The first term on the RHS of this last equality is a U-statistic of order 2. By the Hoeffding decomposition, we get that

$$\frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}} f_n(W_i, W_j) \;=\; f_n(P,P) + \frac{1}{n}\sum_{i=1}^{n}\left[f_n(W_i,P) + f_n(P,W_i) - 2f_n(P,P)\right] \tag{52}$$

$$+\;\frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}}\left[f_n(W_i,W_j) - f_n(W_i,P) - f_n(P,W_j) + f_n(P,P)\right]\,. \tag{53}$$

The term in (53) is a degenerate U-statistic of order 2 having order $O_p(1/n\sigma_n\tau_n)$ as $n \to \infty$. Provided $\sigma_n\tau_n \gg n^{-1/2}$, this term has order $o_p(1/\sqrt{n})$ and so can be ignored. Consider (52). Note that $f_n(W_i, P) = h(U_i)\mathbb{E}G_n(\beta(U_i) \mid U_i)$. It follows from (11) that both $f_n(W_i, P)$ and $f_n(P, P)$ have order $o_p(1/\sqrt{n})$ as $n \to \infty$. Deduce that the only term in the last expression that makes a contribution to the first order asymptotic behavior of $\hat{\beta}_\kappa$ is

$$\frac{1}{n}\sum_{i=1}^{n}\left[f_n(P,W_i) - f_n(P,P)\right]\,. \tag{54}$$

For convenience, we will write this term as

$$\frac{1}{n}\sum_{j=1}^{n}\left[f_n(P,W_j) - f_n(P,P)\right]\,. \tag{55}$$

In order to evaluate $f_n(P, W_j)$, we will fix W_j in $f_n(W_i, W_j)$ and then average over W_i in 2 steps: (i) average over U_i given $X_j\beta(U_i)$ and (ii) average over $X_j\beta(U_i)$. Step (i) involves a change of variable argument and a rescaling by τ_n. Step (ii) involves a change of variable argument and a rescaling by σ_n. As before, we get that the term in (55) is an average of zero-mean iid random vectors with finite variance. A standard CLT shows this term to be $\sqrt{n}$-consistent and asymptotically normally distributed. This proves Lemma 4. $\square$

LEMMA 5 (THE TERM IN P(11)). *If assumptions A1 through A19 hold, then*

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(\hat{U}_i^*)(\hat{U}_i - U_i)\tau_\kappa(\mathcal{X}_i,\mathcal{Z}_i) = \frac{1}{n}\sum_{i=1}^{n} f_n^{(4)}(W_i) + o_p(1/\sqrt{n})$$

as $n \to \infty$ where $f_n^{(4)}(W_i) = f_n(P, W_i) + f_n(P, P, W_i)$ with $f_n(W_i, W_j)$ and $f_n(W_i, W_j, W_k)$ defined in P(20) and P(21), respectively.

PROOF. To save space, we suppress $\tau_\kappa(\mathcal{X}_i,\mathcal{Z}_i)$. We get that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\hat{\delta}(\hat{U}_i^*)(\hat{U}_i - U_i) &= \frac{1}{n}\sum_{i=1}^{n}\delta(U_i)(\hat{U}_i - U_i) &(56)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\delta}(\hat{U}_i^*) - \delta(\hat{U}_i^*)\right](\hat{U}_i - U_i) &(57)\\
&+ \frac{1}{n}\sum_{i=1}^{n}\left[\delta(\hat{U}_i^*) - \delta(U_i)\right](\hat{U}_i - U_i)\,. &(58)
\end{aligned}$$

We will show that the first term on the RHS is $\sqrt{n}$-consistent and asymptotically normally distributed, while the remaining two terms have order $o_p(1/\sqrt{n})$ and so can be neglected.

We start by analyzing the expression in (56). As in the proof of Lemma 4, averages associated with the first two terms in (30) lead to nondegenerate first order asymptotic contributions. Averages associated with the remaining terms make degenerate contributions and are ignored. We analyze the first of the two averages that make nondegenerate contributions. The analysis of the second such term is very similar and so is omitted.

We replace $\hat{U}_i - U_i$ in (56) with

$$\frac{1}{n\alpha_n}\sum_{j=1}^{n}[\{\mathcal{X}_j \le \mathcal{X}_i\}\mathcal{K}_n(\mathcal{Z}_j - \mathcal{Z}_i)\{D_j = D_i\} - \mathbb{E}_i\{\mathcal{X}_j \le \mathcal{X}_i\}\mathcal{K}_n(\mathcal{Z}_j - \mathcal{Z}_i)\{D_j = D_i\}]/f(\mathcal{Z}_i, D_i)\,. \quad (59)$$

Substitute (59) into (56), then combine sums to get that the expression in (56) equals

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\delta(U_i)\left[\{\mathcal{X}_j \le \mathcal{X}_i\}\mathcal{K}_n(\mathcal{Z}_j - \mathcal{Z}_i)\{D_j = D_i\} - \mathbb{E}_i\{\mathcal{X}_j \le \mathcal{X}_i\}\mathcal{K}_n(\mathcal{Z}_j - \mathcal{Z}_i)\{D_j = D_i\}\right]/\alpha_n f(\mathcal{Z}_i, D_i)\,. \quad (60)$$

As before, we may neglect the diagonal terms and take the term in (56) to equal the zero-mean second order U-statistic

$$\frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}}\delta(U_i)\left[\{\mathcal{X}_j\le\mathcal{X}_i\}\mathcal{K}_n(\mathcal{Z}_j-\mathcal{Z}_i)\{D_j=D_i\}-I\!E_i\{\mathcal{X}_j\le\mathcal{X}_i\}\mathcal{K}_n(\mathcal{Z}_j-\mathcal{Z}_i)\{D_j=D_i\}\right]/\alpha_n f(\mathcal{Z}_i,D_i)\,. \tag{61}$$

Define $f_n(W_i,W_j)$ to be equal to the (i,j)th summand in (61). Note that $f_n(W_i,P)=f_n(P,P)=0$. By the Hoeffding composition,

$$\frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}}f_n(W_i,W_j)=\frac{1}{n}\sum_{i=1}^{n}f_n(P,W_i)+\frac{1}{n_{(2)}}\sum_{\boldsymbol{i_2}}\left[f_n(W_i,W_j)-f_n(P,W_j)\right]\,. \tag{62}$$

Standard U-statistic results show that the second term on the RHS of (62) has order $O_p(1/n\alpha_n)$ as $n\to\infty$. This term has order $o_p(1/\sqrt{n})$ provided $\alpha_n\gg n^{-1/2}$. The usual change of variable argument shows that the first term on the RHS of (62) is an average of zero-mean iid random vectors with finite variance. A standard CLT shows that this term is $\sqrt{n}$-consistent and asymptotically normally distributed.

Next, we analyze (58). Let $\gamma(u)$ denote the partial derivative of $\delta(u)$ with respect to U. By a Taylor expansion of each $\delta(\hat{U}_i^*)$ about U_i, we have that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left[\delta(\hat{U}_i^*)-\delta(U_i)\right](\hat{U}_i-U_i) &= \frac{1}{n}\sum_{i=1}^{n}\gamma(\hat{U}_i^{**})(\hat{U}_i^*-U_i)(\hat{U}_i-U_i) &(63)\\ &= \frac{1}{n}\sum_{i=1}^{n}\gamma(\hat{U}_i^{**})(U_i^*-\hat{U}_i)(U_i-\hat{U}_i)\,. &(64)\end{aligned}$$

where, for each i, $\hat{U}_i^{**}$ is between U_i and $\hat{U}_i^*$. Since $\gamma(\cdot)$ is a continuous function on the compact set U_κ, $\gamma(u)$ is uniformly bounded over U_κ. It follows from this and Lemma 2 that the expression in (58) has order $O_p(1/n\alpha_n^2)$. Thus, this term has order $o_p(1/\sqrt{n})$ as $n\to\infty$ provided $\alpha_n^2\gg n^{-1/2}$.

Finally, we analyze the expression in (57). Recall the definition of $\gamma(u)$ above. Let $\hat{\gamma}(u)$ denote the partial derivative of $\hat{\delta}(u)$ with respect to U. By a Taylor expansion of $\hat{\delta}(\hat{U}_i^*)-\delta(\hat{U}_i^*)$ about U_i we get that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left[\hat{\delta}(\hat{U}_i^*)-\delta(\hat{U}_i^*)\right](\hat{U}_i-U_i) &= \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\delta}(U_i)-\delta(U_i)\right](\hat{U}_i-U_i) &(65)\\ &+ \frac{1}{n}\sum_{i=1}^{n}\left[\hat{\gamma}(\hat{U}_i^{**})-\gamma(\hat{U}_i^{**})\right](U_i-\hat{U}_i^*)(U_i-\hat{U}_i) &(66)\end{aligned}$$

where $\hat{U}_i^{**}$ is between $\hat{U}_i^*$ and U_i. Start with (66). Just as integrating kernel functions with respect to u results in a rescaling by a factor of τ_n, differentiating kernel functions with respect to u results in a rescaling by a factor of τ_n^{-1}. This principle can be applied together with (15), (24), and (26) to get that uniformly over i, as $n\to\infty$, $\hat{\gamma}(\hat{U}_i^{**})-\gamma(\hat{U}_i^{**})$ has order $O_p(1/\sqrt{n}\alpha_n\sigma_n\tau_n^4)$. Combine this result with Lemma 2 to see that the term in (66), uniformly over i, as $n\to\infty$, has order $O_p(1/n^{3/2}\alpha_n^3\sigma_n\tau_n^4)$. Deduce that this term has uniform asymptotic order $o_p(1/\sqrt{n})$ provided $\alpha_n^3\sigma_n\tau_n^4\gg n^{-1}$.

We now analyze the term in (65). Differentiate both sides of (15) with respect to u applying the product rule to get that

$$\begin{aligned}\hat{\delta}(u)-\delta(u) &= \frac{\partial}{\partial u}\left[-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}\hat{G}_n(\beta(u)\mid u)\right] \\ &= -[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}\frac{\partial}{\partial u}\left[\hat{G}_n(\beta(u)\mid u)\right] \qquad (67)\\ &+ \frac{\partial}{\partial u}\left[-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}\right]\hat{G}_n(\beta(u)\mid u). \qquad (68)\end{aligned}$$

Start with (68). Focus first on $-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}$. Recall (15). Since the LHS of (15) is continuously differentiable in u and $\hat{G}_n(\beta(u)\mid u)$ is continuously differentiable in u, it follows that $-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}$ is continuously differentiable in u. Thus, for each fixed n, $\frac{\partial}{\partial u}\left[-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}\right]$ is continuous in u on the compact set U_κ and so is bounded. Deduce from this together with (24) and the fact that $-[H(\beta(u)\mid u)]^{-1}$ does not depend on n and is bounded on U_κ, that $-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}=O_p(1)$ uniformly over u as $n\to\infty$. This, together with (26) imply that the term in (68) has order $O_p(1/\sqrt{n}\alpha_n\sigma_n\tau_n^2)$ uniformly over u as $n\to\infty$. Combine this with Lemma 2 and (65) to see that the contribution of (68) to (57) is $O_p(1/n\alpha_n^2\sigma_n\tau_n^2)$ as $n\to\infty$. Provided $\alpha_n^2\sigma_n\tau_n^2\gg n^{-1/2}$, this contribution has order $o_p(1/\sqrt{n})$ as $n\to\infty$.

Finally, consider (67). Argue as in the previous paragraph to see that $-[\hat{H}_n(\hat{\beta}^*(u)\mid u)]^{-1}=O_p(1)$ uniformly over u as $n\to\infty$. Note that by (17),

$$\begin{aligned}\frac{\partial}{\partial u}\left[\hat{G}_n(\beta(u)\mid u)\right] &= \left[\frac{\partial}{\partial u}G_n(\beta(u)\mid u)-\mathbb{E}\frac{\partial}{\partial u}G_n(\beta(u)\mid u)\right]+\frac{\partial}{\partial u}\mathbb{E}G_n(\beta(u)\mid u) \qquad (69)\\ &+ \frac{\partial}{\partial u}\left[\frac{1}{n\sigma_n\tau_n}\sum_{j=1}^n\Lambda_n(\hat{U}_j^*,u)(\hat{U}_j-U_j)\tau_\kappa(\mathcal{Z}_j)\right] \qquad (70)\end{aligned}$$

where for the middle term in (69) we have used the fact that integration and differentiation can be interchanged. By (10) and the fact that differentiation results in a rescaling by τ_n^{-1}, we get that the term in brackets on the RHS of (69) has order $O_p(1/\sqrt{n}\sigma_n\tau_n^2)$ uniformly over u as $n\to\infty$. By (11) and the fact that differentiation results in a rescaling by τ_n^{-1}, we get that the second term on the RHS of (69) has order $o_p(1/\sqrt{n}\tau_n)$ uniformly over u as $n\to\infty$. These facts and Lemma 2 imply that the contribution of the term in (69) to (57) is $O_p(1/n\alpha_n\sigma_n\tau_n^2)+o_p(1/n\alpha_n\tau_n)=O_p(1/n\alpha_n\sigma_n\tau_n^2)$ as $n\to\infty$. Provided $\alpha_n\sigma_n\tau_n^2\gg n^{-1/2}$, this contribution has order $o_p(1/\sqrt{n})$ as $n\to\infty$. Now consider the term in (70). By (17), (18), and Lemma 2, and the fact that differentiation results in a rescaling by τ_n^{-1}, we get that this term has order $O_p(1/\sqrt{n}\alpha_n\sigma_n\tau_n^3)$ as $n\to\infty$. This fact and another application of Lemma 2 imply that the contribution of the term in (70) to (57) is $O_p(1/n\alpha_n^2\sigma_n\tau_n^3)$ as $n\to\infty$. Provided $\alpha_n^2\sigma_n\tau_n^3\gg n^{-1/2}$, this contribution has order $o_p(1/\sqrt{n})$ as $n\to\infty$. This proves Lemma 5.□

Recall the definitions of $f_n^{(j)}(W_i)$, $j=1,2,3,4$ given just prior to the statement of Theorem 1 in the main text.

THEOREM 1. ($\sqrt{n}$-Consistency and Asymptotic Normality) *Let $e=c=1$ with k and* *m arbitrary. If A1 through A19 hold, then, as $n \to \infty$,*

$$\sqrt{n}(\hat{\beta}_\kappa - \beta_\kappa) \rightsquigarrow N(0, \Sigma)$$

where $\Sigma = I\!\!E f_n(W_i) f_n(W_i)'$ with $f_n(W_i) = f^{(1)}(W_i) + f_n^{(2)}(W_i) + f_n^{(3)}(W_i) + f_n^{(4)}(W_i)$.

Proof. Put everything together. Apply a standard CLT for the second term in P(9) together with Lemma 4 and Lemma 5 to get that

$$\hat{\beta}_\kappa - \beta_\kappa = \frac{1}{n} \sum_{i=1}^{n} f_n(W_i) + o_p(1/\sqrt{n})$$

where $f_n(W_i) = f^{(1)}(W_i) + f_n^{(2)}(W_i) + f_n^{(3)}(W_i) + f_n^{(4)}(W_i)$. This proves Theorem 1. □

Appendix B: Trimming and Local Polynomial Estimation

This appendix explains how our trimming scheme prevents problems with boundary bias and ratio bias. It also explains why we choose a trimmed mean, rather than the mean of B, as the estimand of the localize-then-average estimation procedure. Finally, we explain why we do not estimate U_i with higher-order local polynomial estimators, despite their ability to automatically prevent boundary bias.

For simplicity, assume that $\mathcal{X}$ and $\mathcal{Z}$ are scalar random variables. The parameter of interest is the trimmed mean

$$\beta_\kappa = I\!\!E \beta(U) \tau_\kappa(\mathcal{X}, \mathcal{Z}) \tag{71}$$

which we estimate with

$$\hat{\beta}_\kappa = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}(\hat{U}_i) \tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i) \tag{72}$$

where, for each $u \in S_U = [0,1]$, $\hat{\beta}(u) = \underset{b \in B_u}{\operatorname{argmax}}\, \hat{S}_n(b \mid u)$ and

$$\hat{S}_n(b \mid u) = \frac{1}{n\tau_n} \sum_{j=1}^{n} (2Y_j - 1) K_n^*(X_j b) K_n(\hat{U}_j - u) \tau_\kappa(\mathcal{Z}_j)\,. \tag{73}$$

Note that there are two trimming functions: $\tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i) = \{|\mathcal{X}_i| \le \kappa\}\{|\mathcal{Z}_i| \le \kappa\}$ and $\tau_\kappa(\mathcal{Z}_j) = \{|\mathcal{Z}_j| \le \kappa\}$. We discuss the role of each in preventing various types of bias.

Start with $\tau_\kappa(\mathcal{X}_i, \mathcal{Z}_i)$. This trimming function prevents problems with boundary bias and ratio bias. Start with boundary bias. The standard kernel regression estimator $\hat{U}_i$ is an asymptotically biased estimator of U_i when U_i is within the bandwidth τ_n of the boundary of $S_U = [0,1]$. We call this boundary bias. It occurs for any fixed kernel used in standard kernel regression. Recall that $U_i = U(\mathcal{X}_i, \mathcal{Z}_i) = I\!\!P\{\mathcal{X} \le \mathcal{X}_i \mid \mathcal{Z} = \mathcal{Z}_i\}$ and $\hat{U}_i = \hat{U}(\mathcal{X}_i, \mathcal{Z}_i) = \hat{I\!\!P}\{\mathcal{X} \le \mathcal{X}_i \mid \mathcal{Z} = \mathcal{Z}_i\}$. Define

$$\mathcal{U} \;=\; \sup_{|x| \le \kappa, |z| \le \kappa} U(x,z) \qquad \hat{\mathcal{U}} = \sup_{|x| \le \kappa, |z| \le \kappa} \hat{U}(x,z)$$

$$\mathcal{L} = \inf_{|x|\le\kappa,|z|\le\kappa} U(x,z) \qquad \hat{\mathcal{L}} = \inf_{|x|\le\kappa,|z|\le\kappa} \hat{U}(x,z)\,.$$

Since the support of $(\mathcal{X},\mathcal{Z})$ is $I\!R^2$ and $\kappa < \infty$, $0 < \mathcal{L} < \mathcal{U} < 1$. By Lemma 2, $\hat{\mathcal{L}}$ converges in probability to $\mathcal{L}$ and $\hat{\mathcal{U}}$ converges in probability to $\mathcal{U}$. It follows that with probability tending to one as $n \to \infty$, $\tau_\kappa(\mathcal{X}_i,\mathcal{Z}_i)$ trims $\hat\beta(\hat U_i)$ when $\hat U_i$ is within τ_n of 0 or $1-\tau_n$ of 1. This guarantees that the only $\hat\beta(\hat U_i)$ values that play a role in the asymptotic analysis of $\hat\beta_\kappa$ are those whose $\hat U_i$ values are at least τ_n from the boundary of S_U where they are not subject to boundary bias.

The factor $\{|\mathcal{Z}_i| \le \kappa\}$ in $\tau_\kappa(\mathcal{X}_i,\mathcal{Z}_i)$ also prevents so-called ratio bias. Consider the term in P(11). This term involves the factors $\hat U_i - U_i$. In analyzing $\hat U_i - U_i$, terms of the form $[f(\mathcal{Z}_i) - \hat f(\mathcal{Z}_i)]/f(\mathcal{Z}_i)$ (and powers thereof) arise, where $\hat f(\mathcal{Z}_i)$ is a kernel density estimator of $f(\mathcal{Z}_i)$, the density of $\mathcal{Z}$ at $\mathcal{Z}_i$. (See, for example, the geometric expansion of $\hat U(x,z)$ in the proof of Lemma 2.) Note that

$$[f(\mathcal{Z}_i) - \hat f(\mathcal{Z}_i)]/f(\mathcal{Z}_i) = [f(\mathcal{Z}_i) - I\!E\hat f(\mathcal{Z}_i)]/f(\mathcal{Z}_i) + [I\!E\hat f(\mathcal{Z}_i) - \hat f(\mathcal{Z}_i)]/f(\mathcal{Z}_i)\,. \tag{74}$$

Conditional on $\mathcal{Z}_i$, the first term in this decomposition is a deterministic bias term and the second term is a zero-mean stochastic average. Both terms cause problems because of the presence of the density value $f(\mathcal{Z}_i)$ in their denominators. A bias reducing kernel of high enough order can make the numerator of the bias term arbitrarily small, but the ratio can still be large when $f(\mathcal{Z}_i)$ is small. The stochastic term can cause even more serious problems. It's numerator cannot be made arbitrarily small, but rather has stochastic order no smaller than $O_p(1/\sqrt{n})$. Its ratio can be very large when $f(\mathcal{Z}_i)$ is small. However, since $\mathcal{Z}$ has support $I\!R$, these problems can only occur when $\mathcal{Z}_i$ lands in one of the tails of the distribution of $\mathcal{Z}$. The trimming factor $\{|\mathcal{Z}_i| \le \kappa\}$ prevents this from happening, by trimming the summand in P(11) when $|\mathcal{Z}_i|$ gets too big. This prevents ratio bias.

Next, consider the trimming function $\tau_\kappa(\mathcal{Z}_j) = \{|\mathcal{Z}_j| \le \kappa\}$ in (73). Asymptotic analysis of P(10) involves Taylor expansions of the $\hat U_j$s about the corresponding U_js and so leads to analyses of the terms $\hat U_j - U_j$. By the same reasoning as given in the last paragraph, the function $\tau_\kappa(\mathcal{Z}_j)$ trims the jth summand in P(10) when $|\mathcal{Z}_j|$ gets too big, thus preventing ratio bias in these terms.

We note that it is not necessary to do fixed trimming. Provided β exists, we can replace the fixed trimming constant κ with κ_n where $\kappa_n \to \infty$ as $n \to \infty$. The speed at which κ_n converges to infinity must be linked to assumptions about the tail behavior of $f(\mathcal{Z})$. However, such trimming implies that the estimand β_{κ_n} converges to β as $n \to \infty$. Practically speaking, the same effect is achieved by choosing a large fixed κ, and so for the sake of simplicity, we do fixed trimming.

While asymptotically negligible trimming of the sort just described is possible, it is not possible, in general, to take β itself as the estimand and still achieve $\sqrt{n}$-consistency. Establishing $\sqrt{n}$-consistency with β as the estimand would require showing that the difference $\beta_{\kappa_n} - \beta$ has order $O(1/\sqrt{n})$. A straightforward calculation shows that this would require that $I\!P\{|\mathcal{Z}| > \kappa_n\} = O(1/\sqrt{n})$. This, in turn, would require that the density of $\mathcal{Z}$ at $\pm\kappa_n$ be converging to zero very rapidly. However, this same density appears in the denominator of the terms in (74). To prevent ratio bias in these terms it is necessary that the density of $\mathcal{Z}$ at $\pm\kappa_n$ be converging to zero very slowly. It is easy to show that in general, these two conflicting demands cannot be met simultaneously. The real culprit is the stochastic term in decomposition (74). Even if the bias term is identically zero, the stochastic term prevents these conflicting demands from being met. It follows that, apart from special cases when $\beta_\kappa = \beta$ for all $\kappa > 0$ or when Z is discrete so that instrument trimming is unnecessary, if we want to achieve a $\sqrt{n}$-consistent estimator, we must live with a trimmed mean of the distribution of B as an estimand. This is true no matter what estimator we use to estimate U_i. For example, this is true even if we were to replace the standard kernel regression estimators of U_i with general local polynomial (LP) estimators (Fan and Gijbels, 1996).

We estimate U_i with the standard kernel regression estimator, also known as the Nadaraya-Watson (NW) estimator. This is a local polynomial estimator where the local polynomial is a constant. While the NW estimator with bias reducing kernels of high enough order can achieve an arbitrary degree of bias reduction on the interior of the support of the localizing variable, it is an asymptotically biased estimator near the boundaries of the support. We trim on $\mathcal{X}_i$ as well as $\mathcal{Z}_i$ in (72) to prevent this bias, as explained above. However, a comparable higher-order local polynomial estimator can achieve the same degree of bias reduction on the interior as well as near or at the boundary of the support. There is no need to trim on $\mathcal{X}_i$ and $\mathcal{Z}_i$ to prevent boundary bias. So why not use the higher-order LP estimator instead of the NW estimator with bias reducing kernels?

We cite two reasons. First, it is not known (to the authors, at least) whether the known pointwise bounds on the bias of LP estimators at the boundaries of the support of the localizing variable are uniform in the localizing variable. This uniformity is needed to show that remainder terms in asymptotic arguments are small in the appropriate sense.

Secondly, even if the uniformity conditions hold, formally establishing $\sqrt{n}$-consistency and asymptotic normality of $\hat{\beta}_\kappa$ when U_i is estimated with a general local polynomial estimator would be extraordinarily complicated. To see why, assume once again for simplicity that U_i is scalar. A local polynomial estimator of U_i of degree p can be written as the weighted average

$$\sum_{a=1}^{n}\{\mathcal{X}_a \le \mathcal{X}_i\} w_a^{(p)} / \sum_{a=1}^{n} w_a^{(p)}$$

where, for $p > 0$, the weights $w_a^{(p)}$ depend on all the $\mathcal{Z}_i$. For any positive integer m, define $s_m = \sum_{a=1}^{n}(\mathcal{Z}_a - \mathcal{Z}_i)^m K_n(\mathcal{Z}_a - \mathcal{Z}_i)$. For simplicity, we suppress the dependence of s_m on n. Consider the cases $p = 0, 1, 2$, corresponding to the NW, local linear, and local quadratic estimators, respectively. It is straightforward (though tedious) to show that

$$\begin{aligned} w_a^{(0)} &= K_n(\mathcal{Z}_a - \mathcal{Z}_i) \\ w_a^{(1)} &= K_n(\mathcal{Z}_a - \mathcal{Z}_i)[s_2 - (\mathcal{Z}_a - \mathcal{Z}_i)s_1] \\ w_a^{(2)} &= K_n(\mathcal{Z}_a - \mathcal{Z}_i)\left[[s_2 s_4 - s_3^2] - (\mathcal{Z}_a - \mathcal{Z}_i)[s_1 s_4 - s_2 s_3] + (\mathcal{Z}_a - \mathcal{Z}_i)^2[s_1 s_3 - s_2^2]\right] . \end{aligned}$$

Recall that the use of the NW estimator of U_i leads to a complicated analysis of U-statistics of orders

2, 3, and 4 in the proof on Lemma 4. Each of these U-statistics is painstakingly analyzed by means of the Hoeffding decomposition to extract its nonnegligible contribution to the asymptotic distribution of $\hat{\beta}_\kappa$. Now consider local linear estimation. The weight $w_a^{(1)}$ for the local linear estimator is itself a sum and would lead to an analysis of U-statistics of orders 3, 4, and 5 in the proof of Lemma 4. The weight $w_a^{(2)}$ for the local quadratic estimator is a double sum and would lead to an analysis of U-statistics of orders 4, 5, and 6. In general, the weight $w_a^{(p)}$ is a sum over p indices and would lead to the analysis of U-statistics of order $p+2$, $p+3$, and $p+4$. And this is only for scalar U_i. The analysis would be far more complicated for vector-valued U_i.

To avoid this added complexity, we use the NW estimator with higher-order bias reducing kernels. By doing so, we achieve the same order of bias reduction on the interior of the support of the localizing variable as we would by using comparable higher-order local polynomial estimators. By trimming on $\mathcal{X}_i$ (as well as $\mathcal{Z}_i$) we eliminate the problems with bias near the boundary of the support of the localizing variable.

Appendix C

In this appendix we present the results of estimating the Rivers-Vuong estimator where we performed a pre-estimation rescaling of the regressors in the model to achieve an estimated coefficient on the drinks variable of nearly unity. This was done by dividing all regressors in the model by the estimated coefficient on drinks in Table 10 and then applying the Rivers-Vuong estimation procedure to the rescaled data.[3] The results appear in Table 10′ below. We see that the rescaled estimates (and their rescaled standard errors) are nearly identical to those in Table 10 after the post-estimation rescaling. The results in Table 10′ are directly comparable to those in Table 11, and yield nearly identical inferences to those given in Section 5.

Table 10′: Rivers-Vuong Estimator

Variable	Estimate	Std. Error	t value	p value
intercept	-1.1876	0.1177	-10.09	0.000
cigarettes	-0.085	0.044	-1.941	0.052
births	-0.113	0.153	0.53	0.60
age	-0.010	0.03	-0.329	0.742

[3]We could not find a way to enforce the unit restriction on the coefficient of the drinks variable using the probit routine in the statistical package R.